Proceedings of ISA-18 Workshop at LREC2022, pages 127–132
Marseille, 20 June 2022

Annotating Japanese Numeral Expressions for a Logical and Pragmatic Inference Dataset

Kana Koyano[1], Hitomi Yanaka[2], Koji Mineshima[3], Daisuke Bekki[1]
[1]Ochanomizu University, [2]The University of Tokyo, [3]Keio University
{2-1-1 Otsuka, Bunkyo-ku; 7-3-1, Hongo, Bunkyo-ku; 2-15-45, Mita, Minato-ku}, Tokyo, Japan
{koyano.kana, bekki}@is.ocha.ac.jp, hyanaka@is.s.u-tokyo.ac.jp, minesima@abelard.flet.keio.ac.jp

Abstract

Numeral expressions in Japanese are characterized by the flexibility of quantifier positions and the variety of numeral suffixes. However, little work has been done to build annotated corpora focusing on these features and datasets for testing the understanding of Japanese numeral expressions. In this study, we build a corpus that annotates each numeral expression in an existing phrase structure-based Japanese treebank with its usage and numeral suffix types. We also construct an inference test set for numerical expressions based on this annotated corpus. In this test set, we particularly pay attention to inferences where the correct label differs between logical entailment and implicature and those contexts such as negations and conditionals where the entailment labels can be reversed. The baseline experiment with Japanese BERT models shows that our inference test set poses challenges for inference involving various types of numeral expressions.

Keywords: numeral expressions, Japanese, natural language inference, entailment, implicature

1. Introduction

For example, the English sentence "There are three students" can be expressed in Japanese at least in the following three ways.

(1) 学生が　3人　いる
Gakusei-ga san-nin iru
student(s)-NOM three-CLS be-PRS
'There are three students.'

(2) 3人の　学生が　いる
San-nin-no gakusei-ga iru
three-CLS-GEN student(s)-NOM be-PRS
'There are three students.'

(3) 3名の　学生が　いる
San-mei-no gakusei-ga iru
three-CLS-GEN student(s)-NOM be-PRS
'There are three students.'

In (1) and (2), "3人" (*three people*) appears in different positions.
In (2) and (3), the suffix (i.e. classifier) for person is different ("3人" and "3名," respectively). The variety of quantifier positions and numeral suffixes is an important feature of the Japanese language. However, little work has been done to build a corpus focusing on these features or a dataset to test the understanding of Japanese numeral expressions.
Natural Language Inference (NLI) is the semantic task of determining whether the hypothesis is true, false, or neither, when the premises are all true. It is considered one of the core knowledge underlying natural language understanding. Recently, not only semantic modes of reasoning, but also pragmatic modes of reasoning have been explored in the field of natural language processing (Jeretic et al., 2020). These two modes of inferences correspond to *entailment* and *implicature*, which have been discussed in the linguistic literature (Levinson, 1983; Horn, 1989; Levinson, 2000). For example, consider the following premise–hypothesis pair.

(4) 男性が　道端に　4人　座って いた
Dansei-ga mitibata-ni yo-nin suwatte ita
men-NOM street-LOC four-CLS sit-PROG be-PST
'Four men were sitting on the street.'

(5) 男性が　道端に　5人　座って いた
Dansei-ga mitibata-ni go-nin suwatte ita
men-NOM street-LOC five-CLS sit-PROG be-PST
'Five men were sitting on the street.'

We use different labels (*logical label* and *pragmatic label*) for the judgments based on entailment and implicature, respectively, because they may differ on the same premise–hypothesis pair: the logical label for this inference is NEUTRAL, whereas the pragmatic label is CONTRADICTION. The latter is so because, along with Grice's maxim of quantity, if the speaker knew that five people were sitting there, they would say so, and the fact that they dared to say (4) implies that there is no fifth person. In other words, in entailment, inferences are made only from the semantic information contained in the premises and hypothesis. In implicature, however, the assumption that normal conversation proceeds according to the *co-operative principle* gives rise to information not semantically included in the utterance, considering the context and the speaker's intention, as suggested in Grice's maxims of conversation (Levinson, 1983; Grice, 1989).
In this study, we construct a Japanese corpus in which numeral expressions are annotated regarding the classi-

fication of syntactic patterns and the usage of numeral expressions. We use sentences containing numeral expressions extracted from the NINJAL Parsed Corpus of Modern Japanese (NPCMJ) (NINJAL, 2016), which is a phrase structure-based treebank for Japanese. Furthermore, we construct an inference test set for numeral expressions based on this corpus, which reports two modes of judgments, entailment and implicature for each premises-hypothesis pair.

In this paper, we report on the design of the corpus and inference test set as well as the evaluation results of a baseline experiment. The constructed dataset will be made publicly available in a research-usable format[1].

2. Related Work

Regarding the study of NLI focusing on English numeral expressions, (Naik et al., 2018) presents an inference dataset that contains 7,596 premise–hypothesis pairs, with 2,532 in each gold label (entailment, neutral, and contradiction). However, a recent study (Liu et al., 2019) has pointed out that the majority of problems (about 82% of the total) in this dataset can be solved using a few heuristic rules, which is due to the fact that the inference of numeral expressions is constructed using a simple template.

Jeretic et al. (2020) provided an English NLI dataset that focuses on the distinction between logical entailment, presupposition, and implicature. It also contains inference problems for scalar implicature triggered by numeral expressions. However, it is automatically constructed from templates and thus the sentences are relatively simple.

Cui et al. (2022) examined the extent to which multilingual pre-trained language models capture the behavior of generalized quantifiers including various types of numeral expressions in English. Their experiments showed that quantifiers cause performance drops for NLI and question answering models. We can say that numeral expressions pose an important challenge to the study of NLI and other tasks for natural language understanding. Our corpus and inference dataset focusing on numeral expressions in Japanese contribute further insight on how pre-trained language models work.

Previous Japanese inference datasets include JSeM (Kawazoe et al., 2017), the formal semantics test set (the Japanese version of FraCaS); JSNLI (Yoshikoshi et al., 2020), the Japanese version of English SNLI (Bowman et al., 2015); JSICK (Yanaka and Mineshima, 2021b), the Japanese version of English SICK(Marelli et al., 2014); and a crowdsourced dataset from real text, reputation, a travel information website (Hayashibe, 2020), and other sources. However, in these datasets, the syntactic and semantic diversity of Japanese numeral expressions is not fully taken into account. Narisawa et al. (2013) investigated cases where numeral expressions are problematic in Japanese NLI and implemented a module for normalizing numeral expressions. They classify premise–hypothesis pairs containing numeral expressions into seven categories and describe the process required to correctly determine the entailment relation, but they do not consider the difference between the two inference types (namely, entailment and implicature), which may give rise to different judgements according to the classification of numeral expressions and numeral suffixes.

Given these considerations, in our study, we first annotate numeral expressions in a Japanese corpus containing real Japanese texts and classify them according to their usages and the difference in numeral suffixes. By using the annotated corpus, we create an inference dataset involving numeral expressions annotated with entailment and implicature labels.

3. Syntax and Semantics of Japanese Numeral Expressions

3.1. Classification of numeral suffixes

According to Iida (2021), numeral suffixes are classified into three categories: sortal suffixes, unit-forming suffixes, and measure suffixes. In addition, some words have an ordinal number suffix (Okutsu, 1996), which expresses order within a time line or sequence. Thus, in this study, we propose a taxonomy that extends the three types of numeral suffixes in Iida (2021) with ordinal number suffix. Examples of each type of numeral suffix are shown in Table 1.

Table 1: Examples and the number of occurrences of each type of numeral suffix

Type	Example	Occurrence
sortal suffixes	人, 頭, 冊, 枚	56
unit-forming suffixes	瓶, 箱, 袋, パック, 切れ	13
measure suffixes	リットル, 円, バイト	74
ordinal number suffixes	月, 日, 番, 位	107

The classification of some numeral suffixes is not uniquely determined by their surface forms but depends on the context and usage. For example, “階” (*floor*) in “会議室は建物の3階にある” (*the conference room is on the third floor of the building*) is an ordinal number suffix, while “階” (*floor*) in “ここから3階のぼったところに会議室がある” (*there is a conference room three floors up from here*) is a measure suffix. The former refers to a specific location of the conference room, while the latter refers to the number of floors to go up. Note that, in the latter, the conference room is not necessarily located on the third floor.

3.2. Position of occurrence of numeral expressions

Encyclopedia of Japanese (EJ) (Yazawa, 1988) classified the syntactic patterns containing numeral expressions into four categories: *Q no NC*, *N no QC*, *NCQ*, and *NQC*, where *Q*, *N*, *C* stand for a numeral together

[1] https://github.com/KanaKoyano/numeral-expressions-corpus

with a classifier, a common noun, and a case marker, respectively. Iwata (2013) added two categories to the classification of EJ, *predicate* type and *De* type. In this study, we extended the classification by adding the following types, the examples of which are shown in Table 2.

***QV*:** *Q* semantically modifies the verb *V*.

***NvCQ*:** *Q* is a predicate on the event noun phrase *Nv*.

***N dropout*:** The so-called pronominal usage in which *no N* of *Q no NC* is omitted.

***QtQ*:** A time expression and a numeral expression are adjacent, such as in "1時間(で)500円" (*500 yen for 1 hour*) and "1ヶ月に1回" (*once a month*).

***idiom*:** Idiomatic and conventional usages.

***(Q)*:** A numeral expression enclosed within a bracket.

Table 2: Example and the number of occurrences of each position of numeral expressions

Type	Example	Occurrences
Q no NC	3人の学生が来た	31
N no QC	学生の3人が来た	11
NCQ	学生が3人来た	53
NQC	学生3人が来た	11
predicate	来た学生は3人だ	1
De	学生が3人で来た	7
QV	東京に3回行った	74
NvCQ	渡米したことは2回ある	6
N dropout	3人はお金を払った	24
QtQ	1時間500円かかる	3
idiom	1人暮らし, 8人兄弟	14
(Q)	（1998年）	15

3.3. Usage of numeral expressions

In addition to the usage of the numeral expression *Q* studied by Iwata (2013), the present study adds three new usage categories of *Q* by modifying the noun *N* and four more usage categories of *Q* by modifying the verb *V*. In addition, we add the usage of the expression *Q* by modifying *Nv* and idiomatic usage. In summary, we classify each numeral expression according to ten usage categories. The usage classifications and their examples are shown in Table 3.

4. Semantic Annotation of Numeral Expressions

In this study, 250 numeral expressions of sentences extracted from the NPCMJ were annotated by a graduate student with a background in linguistics.

Table 3: Example and the number of occurrences of each usage of numeral expression

Type	Example	Occurrence
Q represents the category information of *N*	3人の学生 (three students)	60
Q represents the number of elements that constitutes *N*	5人の集団 (a group of 5 people)	8
Q represents part of the elements that constitutes *N*	集団の1人 (one person from the group)	7
Q represents an attribute or characteristic of *N*	50歳の男性 (50 years old man)	64
Q for the number of times *V* has taken place	2回来る (come twice)	1
Q for the period in which *V* took place	3日滞在する (stay for 3 days)	21
Q representing the time that *V* took place	9時に来る (come at 9:00)	57
Q for characteristics of *V*	2%上昇する (increase by 2%)	13
Q to qualify *Nv*	渡航歴は2回 (two trips)	5
idiom	1人暮らし (living alone)	14

Table 4: Occurrences of upward/downward monotone inference

upward monotone	downward monotone
1173	118

Semantic annotation We assigned <num> tags to the numeral expressions that appeared in sentences, and made annotations for the classification of numeral suffixes, position of occurrence, and usage of numeral expression, as described in Section 3. When multiple numeral expressions appeared in a sentence, we marked the target expression with the <num> tag. The number of occurrences of each type of numeral suffixes, each position of numeral expression, and each usage in the corpus are shown in Table 1, Table 2, and Table 3, respectively.

5. Inference Test Set of Numeral Expressions

5.1. Data creation

We create an inference test set from a corpus of numeral expressions. We use each sentence in the corpus for a premise sentence T. The hypothesis sentence was created using the sentence annotated in Section 4. We select the clause that does not change the meaning of the numeral expression tagged with <num> as in (6), change the numeral, and add a quantifier modifier, as in (7).

(6) 仙台都市圏（広域行政圏）の
Sendai-tosi-ken（Kouiki-gyousei-ken）-no
推計人口は　約<num>151万人</num>
suikei-zinkoo-wa yaku-151man-nin
で…
de…
Sendai-metropolitan-area (greater-administrative-area)-GEN estimated-population-NOM approximately-1.51-million-CLS be-cont
'The estimated population of the Sendai metropolitan area (greater administrative area)

Table 5: Examples of the inference test set

premise T and hypotheses H_- and H_+	gold labels (T, H_-) logical	(T, H_-) pragmatic	(T, H_+) logical	(T, H_+) pragmatic
T: 前回１９９７年の税率アップ時を参考にすれば、昨年１２月～３月の駆け込み需要で前年比１％の売り上げ増が見込まれる半面、ことし４月以降は４～５％程度の落ち込みが予想される (If the previous tax rate increase in 1997 is used as a reference, a 1% year-on-year increase in sales is expected from December to March of last year due to rush demand, while a 4-5% decline is expected from April of this year.) H_-: 前回の税率アップは１９９６年より後だった (The last tax rate increase was later than 1996.) H_+: 前回の税率アップは１９９８年より後だった (The last tax rate increase was later than 1998.)	ENTAILMENT	ENTAILMENT	CONTRADICTION	CONTRADICTION
T: 勿論、私ひとりで四升呑みほしたわけでは無い (Of course, I didn't finish all four bottles by myself.) H_-: 勿論、私ひとりで三升以上呑みほしたわけでは無い (Of course, I didn't finish more than three bottles by myself.) H_+: 勿論、私ひとりで五升以上呑みほしたわけでは無い (Of course, I didn't finish more than five bottles by myself.)	NEUTRAL	NEUTRAL	ENTAILMENT	ENTAILMENT
T: あの頃は、100ドルを円に両替すれば、12000円になりました (Back then, if you exchanged $100 into yen, it became 12,000 yen.) H_-: あの頃は、50ドル以上を円に両替すれば、12000円になりました (Back then, if you exchanged more than $50 into yen, it became 12,000 yen.) H_+: あの頃は、150ドル以上を円に両替すれば、12000円になりました (Back then, if you exchanged more than $150 into yen, it became 12,000 yen.)	NEUTRAL	NEUTRAL	ENTAILMENT	ENTAILMENT

Table 6: Results of baseline experiments using Japanese BERT (accuracies of correct responses)

training	logical label all	ENTAILMENT	CONTRADICTION	NEUTRAL	pragmatic label all	ENTAILMENT	CONTRADICTION	NEUTRAL
JSICK	32.22%	70.83%	10.61%	17.74%	30.83%	70.90%	9.62%	16.67%
JSNLI	41.21%	70.83%	35.52%	5.66%	44.46%	70.67%	35.23%	6.67%

Table 7: Inference test set statistics

	ENTAILMENT	CONTRADICTION	NEUTRAL
logical label	432	594	165
pragmatic label	433	738	120

is approximately 1.51 million, and . . . '

(7) 仙台都市圏の　　推計人口は
Sendai-tosi-ken-no suikei-zinkoo-wa
160万人　以上 である
160man-nin izyoo dearu
Sendai-metropolitan-area-GEN estimated-population-NOM 1.6-million-CLS over be-PRS
'The estimated population of the Sendai metropolitan area is over 1.6 million'

We rephrase numerals in a premise sentence T with both a lower number (H_-) and a higher number (H_+) and create two premise–hypothesis pairs (T, H_-) and (T, H_+) from T.

As in (6), when a modifier such as "約" (*approximately*) is added to a numeral expression, all judgment labels would become NEUTRAL if the hypothesis sentence is created with too small a number. In such cases, the numbers in a hypothesis sentence were modified so that the pragmatic label becomes as CONTRADICTION while preserving its naturalness. In cases where adding a modifier would result in making an unnatural sentence as in (7), we changed the word order of a sentence while maintaining its original meaning in creating a hypothesis sentence.

In this study, we did not use sentences involving idiomatic usage because changing the number or adding a modifier of such sentences would make the rephrased sentences unnatural.

5.2. Monotonicity inference

We also create inference problems involving the so-called monotonicity inference triggered by numeral expressions. If M is a more specific concept (subconcept) of N, then a sentence $\varphi(M)$ involving M usually entails a sentence $\varphi(N)$ involving N. We call such inference *upward monotone* inference. In the case of numeral expressions, for example, "200人" (*200 people*) refers to a subconcept of "100人" (*100 people*), so if the sentence *There are 200 people in the hall* is true, then the sentence *There are 100 people in the hall* is also true. However, if numeral expressions are embedded in *downward monotonic* contexts such as negations and conditionals, the entailment relation is inverted. Here a sentence containing the more general concept $\varphi(N)$ entails a sentence containing a more specific concept $\varphi(M)$. For example, the sentence *There were not 100 people in the hall* entails the sentence *There were not 200 people in the hall*.

The first example in Table 5 is a premise-hypothesis pair in an upward monotone context. The second and third examples are premise-hypothesis pairs in a downward monotone context involving negation and conditionals, respectively. Table 4 shows the number of occurrences of upward and downward monotone inference. At present, the number of downward monotone inference is small, reflecting the fact that expressions that trigger this type of inference is rare in the corpus. It is left for future work to annotate more examples of downward monotone inferences involving numeral expressions.

5.3. Inference test set

The inference test set created in this study contains 1,291 premise–hypothesis pairs. One annotator assigned logical (entailment) and pragmatic (implicature) labels to each pair in the inference test set.

The statistics of the inference test set are shown in Table 7 and examples of premise and hypothesis sentences are shown in Table 5. We can see that the numbers of CONTRADICTION and NEUTRAL judgments for logical and pragmatic labels are different because some of those that are NEUTRAL for logical labels are CONTRADICTION for pragmatic labels.

5.4. Baseline experiments

To evaluate the extent to which current standard pretrained language models can handle inferences that require an understanding of numeral expressions, we conducted an evaluation experiment using Japanese BERT (Devlin et al., 2019) as a baseline model. In the experiment, we used two standard Japanese NLI datasets to finetune BERT models on the NLI task: Japanese SICK datasets (JSICK, 5,000 pairs) (Yanaka and Mineshima, 2021a) and Japanese SNLI datasets (JSNLI, 530,000 pairs) (Yoshikoshi et al., 2020).

Table 6 shows the evaluation results of the NLI model. Overall, the accuracies to the Japanese BERT tend to be higher for models trained on JSNLI than for those trained on JSICK, but both were below 50%. In particular, the accuracy for ENTAILMENT was over 60%, while the accuracies for CONTRADICTION and NEUTRAL were both below 40%, suggesting a tendency to predict ENTAILMENT when the model is trained on an existing dataset. As for the difference in training data, the accuracy for CONTRADICTION was higher for both logical label and pragmatic label when JSNLI was used than when JSICK was used, which might be due to the larger number of training data used for JSNLI.

Table 8 shows the accuracies for each position of occurrence of the numeral expressions. The results show that the performance on inference examples involving numeral expressions of De types was low. One possible reason for the low performance is that numeral expressions of De types might be not frequently appear in general, including the training data. Thus models struggled with predicting correct labels for inferences involving numeral expressions of De types.

Table 8: Accuracies for each position of occurrence

Type	logical label		pragmatic label	
	JSICK	JSNLI	JSICK	JSNLI
Q no NC	29.70%	35.76%	24.85%	36.97%
N no QC	28.79%	40.91%	25.76%	39.39%
NCQ	32.00%	44.33%	31.00%	48.33%
NQC	31.82%	48.48%	33.33%	53.03%
predicate	0.00%	25.00%	0.00%	25.00%
De	28.21%	43.59%	23.08%	41.03%
QV	32.39%	39.95%	33.10%	44.68%
NvCQ	27.27%	59.09%	27.27%	45.45%
N dropout	39.23%	38.46%	36.92%	45.38%
QtQ	33.33%	33.33%	20.00%	40.00%
(Q)	34.43%	42.62%	31.15%	42.62%

6. Conclusion

In this study, we constructed a Japanese corpus of numeral expressions as well as semantic annotations including the classification of numeral suffixes and their usage. We also created a logical and pragmatic inference test set from the corpus of numeral expressions. As a baseline experiment, we evaluated Japanese BERT on our inference test set. The experiment showed that our inference test set for numeral expressions constructed is challenging enough for the current standard NLI models. When constructing the annotated corpus for numeral expressions and the inference dataset, we focused on the characteristics of Japanese, such as the flexibility of quantifier positions and the diversity of numeral suffixes. Future work remains to annotate and analyze more semantically complex phenomena, i.e., those phenomena that have been studied in the previous analysis of quantification in English (Bunt, 2020), including the scope of quantification, definiteness, and the distributive/collective distinction in Japanese numeral expressions. We will also continue to expand our numeral expression corpus and inference dataset as well as analyze the current NLI models on our inference dataset.

Acknowledgments

We thank the five anonymous reviewers for their helpful comments and feedback. This work was partially supported by JST CREST Grant Number JPMJCR20D2, Japan, and JST PRESTO Grant Number JPMJPR21C8, Japan.

References

Bowman, S. R., Angeli, G., Potts, C., and Manning, C. D. (2015). A large annotated corpus for learning natural language inference. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 632–642.

Bunt, H. (2020). Annotation of quantification: The current state of ISO 24617-12. In *Proceedings of the 16th Joint ACL - ISO Workshop on Interoperable Semantic Annotation*, pages 1–12, Marseille, May. European Language Resources Association.

Cui, R., Hershcovich, D., and Søgaard, A. (2022). Generalized quantifiers as a source of error in multilingual NLU benchmarks. In *Proc. of NAACL*.

Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*.

Grice, P. (1989). *Studies in the Way of Words*. Harvard University Press.

Hayashibe, Y. (2020). Japanese realistic textual entailment corpus. In *Proceedings of the 12th Language Resources and Evaluation Conference*, pages 6827–6834.

Horn, L. R. (1989). *A Natural History of Negation*. University of Chicago Press.

Iida, T. (2021). Japanese semantics and the mass/count distinction. In Chungmin Lee, et al., editors, *Numeral Classifiers and Classifier Languages*. Routledge, London.

Iwata, K. (2013). *Nihongo Suuryo Hyoogen no Shoso (Aspects of Japanese Numeral Expressions)*. Kuroshio.

Jeretic, P., Warstadt, A., Bhooshan, S., and Williams, A. (2020). Are natural language inference models IMPPRESsive? Learning IMPlicature and PRESupposition. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 8690–8705, Online, July. Association for Computational Linguistics.

Kawazoe, A., Tanaka, R., Mineshima, K., and Bekki, D. (2017). An inference problem set for evaluating semantic theories and semantic processing systems for japanese. In *New Frontiers in Artificial Intelligence*, pages 58–65.

Levinson, S. C. (1983). *Pragmatics*. Cambridge University Press.

Levinson, S. C. (2000). *Presumptive Meanings: The Theory of Generalized Conversational Implicature*. MIT press.

Liu, N. F., Schwartz, R., and Smith, N. A. (2019). Inoculation by fine-tuning: A method for analyzing challenge datasets. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 2171–2179, Minneapolis, Minnesota, June. Association for Computational Linguistics.

Marelli, M., Menini, S., Baroni, M., Bentivogli, L., Bernardi, R., and Zamparelli, R. (2014). A SICK cure for the evaluation of compositional distributional semantic models. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, pages 216–223.

Naik, A., Ravichander, A., Sadeh, N., Rose, C., and Neubig, G. (2018). Stress test evaluation for natural language inference. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 2340–2353, Santa Fe, New Mexico, USA, August. Association for Computational Linguistics.

Narisawa, K., Watanabe, Y., Mizuno, J., Okazaki, N., and Inui, K. (2013). Is a 204 cm man tall or small? acquisition of numerical common sense from the web. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 382–391, Sofia, Bulgaria, August. Association for Computational Linguistics.

Okutsu, K. (1996). *Syuui Nihon-Bupoo-Ron*. Hitsuji Shobo.

Yanaka, H. and Mineshima, K. (2021a). Assessing the generalization capacity of pre-trained language models through Japanese adversarial natural language inference. In *Proceedings of the Fourth BlackboxNLP Workshop on Analyzing and Interpreting Neural Networks for NLP*, pages 337–349, Punta Cana, Dominican Republic, November. Association for Computational Linguistics.

Yanaka, H. and Mineshima, K. (2021b). JSICK: Japanese constructive inference and similarity dataset construction (in Japanese). In *The Thirty-fifth Annual Meeting of Japanese Society for Artificial Intelligence*.

Yazawa, M. (1988). Suuryo no Hyougen (Expression of quantity). In Haruhiko Kindaichi, et al., editors, *Nihongo Hyakka Daijiten (Encyclopedia of Japanese)*. Taishukan Shobo.

Yoshikoshi, T., Kawahara, D., and Sadao, K. (2020). Multilingualization of natural language inference datasets using machine translation (in Japanese). In *The 244th Meeting of Natural Language Processing*.

7. Language Resource References

NINJAL. (2016). *NINJAL Parsed Corpus of Modern Japanese. (Version 1.0)*.